SPEAKER VERIFICATION APPARATUS AND METHOD

BACKGROUND OF THE INVENTION

1. Field of the Invention

5   The present invention relates to a speaker verification apparatus and method for determining by the voice of the speaker whether or not the speaker is an authorized user based on the feature parameters of the voices that are previously registered.

10   2. Description of the Prior Art

In recent years, with the development of computer technologies, a communication environment has been developed rapidly.   With the development of such a communication environment, computer telephony integration through the telephone has become common in ordinary
15   homes.

In the field of such computer telephony integration through the telephone, a problem may arise when accessing information that should not be known to people other than the authorized person or a specific group of authorized people, such as private information or information
20   subjected to secrecy obligation.   More specifically, for example, when a push-button telephone is used, it is possible to acquire an access authority to information by inputting a password by an operation of pushing buttons of the telephone.   However, when the password is known to unauthorized people, they can access the information easily
25   although they are not duly authorized.   For this reason, there is a need of verifying whether or not the person who tries to access the information is the duly authorized person or one of a specific group of authorized people using the voice, which is inherent to the individual.   In order to ensure such a security function, it is important that the registration of
30   voices for verification or the determination of the threshold for judging whether or not the input voice is the voice of an authorized person does not cause an excessive burden to the user.

Conventionally in general, a fixed and predetermined value has been used as a threshold for determining whether or not the speaker is an authorized person.   More specifically, as shown in Fig. 1, a verification distance between an input voice and a previously registered voice is

5 calculated and compared to a predetermined threshold.   When the verification distance is equal to or shorter than the predetermined threshold ("−" in Fig. 1), it is determined that the speaker is an authorized person.   When the verification distance is longer than the predetermined threshold ("+" in Fig. 1), it is determined that the speaker

10 is an unauthorized person.

It is desirable that such a threshold is set to a value as described below.   In Fig. 2, FR (false rejection error rate), which is a probability of the case where the determination that the speaker should be rejected as an unauthorized person is erroneous, is plotted in the vertical axis

15 against the threshold of the verification distance in the horizontal axis. Similarly, FA (false acceptance error rate), which is a probability of the case where an unauthorized person is erroneously accepted, is plotted in the vertical axis against the threshold of the verification distance in the horizontal axis.   When the threshold is a small value, the rate FA of

20 erroneous acceptance of an unauthorized person is low, whereas the rate FR of erroneous rejection of an authorized person is high.   On the other hand, when the threshold is a large value, the rate FR of erroneous rejection of an authorized person is low, whereas the rate FA of erroneous acceptance of an unauthorized person is high.   Therefore, it is desirable

25 to set the threshold to be an appropriate value depending on the level of importance of the two error rates.   It is general to perform verification using a value that allows the two error rates to be eventually equal experimentally as the threshold.

However, in the above-described method, it is necessary to be

30 aware of the tendency of the false rejection error rate FR and the false acceptance error rate FA beforehand to set the threshold.   However, it is difficult to know the two error rates before being used.   Therefore, a

2

preliminary experiment is performed to seek an approximate value, or the threshold is updated whenever it is required at the time of using the system. The method of performing a preliminary experiment is disadvantageous for the following reasons. Because of the difference in the conditions between when the preliminary experiment is performed and when the system is actually used, it is often necessary to perform a test again when using the system. In addition, in order to obtain the false rejection error rate FR, it is necessary for an authorized person (user) to give his/her voice many times, which causes a large burden to the user and is unpractical. On the other hand, the method of updating the threshold whenever it is required at the time of using the system is disadvantageous because updating the threshold causes a large burden to the user as well.

Furthermore, the voice of an authorized person can change over time, and in general, accurate identification of the speaker is difficult when noise such as background sound is mixed therewith.

SUMMARY OF THE INVENTION

Therefore, with the foregoing in mind, it is an object of the present invention to provide a speaker verification apparatus and method whose implementation environment can be set without an excessive burden to a user and that can specify the speaker in a high accuracy.

A speaker verification apparatus of the present invention includes an identity claim input part to which an identity claim is input; a speaker selecting part for selecting the voice information of the registered speaker corresponding to the identity claim input to the identity claim input part; a speaker storing part for storing voice information of speakers; a voice input part to which a voice of a speaker is input; a voice analyzing part for analyzing the voice input to the voice input part; a speaker distance calculating part for calculating a verification distance between a feature parameter of the input voice and that of the voice of the registered speaker and the speaker distances between a feature parameter of the

3

input voice and those of the voices of speakers other than the registered speaker that are stored in the speaker sorting part, based on the analysis results of the voice analyzing part and the voice information stored in the speaker storing part; and a speaker judging part for determining whether or not the input voice matches the registered speaker corresponding to the input identity claim. The speaker verification apparatus further includes a false acceptance error rate input part to which a false acceptance error rate is input as a threshold, the false acceptance error rate being predetermined by a system manager or a user or adjustable depending on the performance, and a distribution estimating part for obtaining a probability distribution of interspeaker distances based on the speaker distances calculated in the speaker distance calculating part. The speaker judging part determines that the input voice is the voice of the person specified by the identity claim, in the case where the verification distance calculated in the speaker distance calculating part is included in a region defined by the input false acceptance error rate in the probability distribution of the interspeaker distances. Herein, "interspeaker distances" means a distance with a speaker template that is not the template of the person specified by the identity claim.

In this embodiment, a fixed threshold of a verification distance is not used, but a probability of the interspeaker distances is calculated each time a system is used, and a threshold is determined based on the false acceptance error rate in the distribution of the interspeaker distances. Therefore, a criterion for judging the speaker closer to the theoretical values of a statistical probability distribution can be obtained. In addition, the false acceptance error rate can be maintained closer to the theoretical values even if the voice input environment changes and a noise is mixed. Thus, the verification accuracy of the speaker verification can be maintained high without being affected by the aging of the input voice. The present invention is based on the empirical fact that although the speaker distance itself constantly changes by external factors such as the difference in the environment where a voice is input

4

and the aging of the voice, the relationship with respect to the interspeaker distances between the input voice and the other registered speakers hardly changes.

Another aspect of the present invention is software that executes the functions of the above-described speaker verification apparatus. More specifically, the present invention is directed to a computer-readable recording medium on which the method for verifying a speaker or steps of the method are recorded as a program. The method includes inputting an identity claim; selecting voice information of a registered speaker corresponding to the input identity claim; inputting a voice of the speaker; analyzing the input voice; calculating a verification distance between the input voice and the voice of the registered speaker and the speaker distances between the input voice and voices of registered speakers other than the registered speaker, based on the analysis results and the voice; and determining whether or not the input voice matches the registered speaker corresponding to the input identity claim. The method further includes inputting a false acceptance error rate as a threshold, the false acceptance error rate being predetermined by a system manager or a user or adjustable depending on the performance; and obtaining a probability distribution of the interspeaker distances based on the calculated speaker distances. It is determined that the input voice is the voice of the person specified by the identity claim, in the case where the calculated verification distance is included in a region defined by the input false acceptance error rate in the probability distribution of the interspeaker distances.

This embodiment can realize a speaker verification apparatus as described below by loading the program onto a computer and executing the program. A fixed threshold of a verification distance is not used, but a probability distribution of the interspeaker distances is calculated each time a system is used, and a threshold is determined based on the false acceptance error rate in the distribution. Therefore, a criterion for judging the speaker closer to the theoretical values of a statistical

5

probability distribution can be obtained.   In addition, the false acceptance error rate can be maintained closer to the theoretical values even if the voice input environment changes and a noise is mixed.   Thus, the verification accuracy of the speaker verification can be maintained high without being affected by the aging of the input voice.

These and other advantages of the present invention will become apparent to those skilled in the art upon reading and understanding the following detailed description with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a general conceptual diagram of speaker verification.

Fig. 2 is a diagram for illustrating a method for specifying a threshold in a conventional speaker verification method.

Fig. 3 is a structural block diagram of a speaker verification apparatus of an embodiment of the present invention.

Fig. 4 is a diagram for illustrating a method for specifying a threshold in the speaker verification apparatus of an embodiment of the present invention.

Fig. 5 is a structural block diagram of a speaker verification apparatus of one example of the present invention when verifying the speaker.

Fig. 6 is a graph showing the experimental results obtained when a speaker verification method of one example of the present invention is used under a quiet environment.

Fig. 7 is a graph showing the experimental results obtained when a speaker verification method of one example of the present invention is used under a noisy environment.

Fig. 8 is a graph showing the experimental results obtained when a speaker verification method of one example of the present invention is used for each utterance period.

Fig. 9 is a structural block diagram of a speaker verification apparatus of one example of the present invention when registering a

6

speaker.

Fig. 10 is a flowchart of the processes for verifying the speaker in the speaker verification apparatus of an embodiment of the present invention.

Fig. 11 is a flowchart of the processes for registering a speaker in the speaker verification apparatus of an embodiment of the present invention.

Fig. 12 is a diagram of an illustrative recording medium.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hereinafter, a speaker verification apparatus of an embodiment of the present invention will be described with reference to the accompanying drawings. Fig. 3 is a structural diagram showing the principle of the speaker verification apparatus of an embodiment of the present invention. Referring to Fig. 3, numeral 31 denotes an identity claim input part. Numeral 32 denotes a speaker template selecting part. Numeral 33 denotes a speaker template storing part. Numeral 34 denotes a voice input part. Numeral 35 denotes a voice analyzing part. Numeral 36A denotes a verification distance calculating part. Numeral 36B denotes a speaker distance calculating part. Numeral 37 denotes a distribution estimating part. Numeral 38 denotes a false acceptance error rate input part. Numeral 39 denotes a speaker judging part.

In Fig. 3, at the time of using a system, an identity claim is input to the ID input part 31. Then, the speaker template selecting part 32 selects a template corresponding to the identity claim from templates of a plurality of speakers that are previously registered in the speaker template storing part 33 and sends the selected template to the verification distance calculating part 36A. At the same time, the templates of the registered speakers other than the speaker corresponding to the identity claim are sent out to the speaker distance calculating part 36B.

Next, in the voice analyzing part 35, a voice input to the voice

7

input part 34 is converted into a feature parameter for speaker verification and sent out to the verification distance calculating part 36A and the speaker distance calculating part 36B. The verification distance calculating part 36A calculates the distance d_{id} between the voice

5   template of the speaker corresponding to the identity claim and the feature parameter of the input voice.

On the other hand, the speaker distance calculating part 36B calculates the distances d_1, d_2, ... and d_N between the voice templates of N other registered speakers and the feature parameter of the input voice

10   and delivers the results to the distribution estimating part 37. The distribution estimating part 37 estimates a probability distribution function F(d) of the speaker distances between the voices of the registered speakers other than the speaker corresponding to the input identity claim and the input voice, using the calculated N distances d_1, d_2, ... and d_N with

15   respect to the other registered speakers and delivers the result to the speaker judging part 39.

The estimation of the probability distribution function F(d) leads to a probability density function f(d). The area of the function shown in the probability density function f(d) indicates a probability value. The

20   relationship between the probability distribution function F(d) and the probability density function f(d) is that as shown in Equation 1.
Equation 1

$$F(x) = \int_{-\infty}^{X} f(t)\, dt$$

25

Therefore, the speaker judging part 39 judges the speaker based on the probability density function f(d) in the following manner. When the speaker distance d_{id} with respect to the speaker corresponding to the identity claim is within the region defined by the level of significance p of

30   regarding an unauthorized person as the person specified by the ID, which is previously designated in the false acceptance error rate input

8

part 38, it is determined that the speaker is the person specified by the ID. When the distance d_{id} is not within the region, it is determined that the speaker is not the person specified by the ID. In the determination based on the probability distribution function $F(d)$, when $F(d_{id}) < p$ is satisfied, the speaker is the person specified by the ID. When $F(d_{id}) \geqq p$ is satisfied, the speaker is not the person specified by the ID.

Fig. 4 shows a diagram illustrating the method for judging the speaker by the speaker judging part 39. In the case where the probability density function $f(d)$ is already obtained, the hatched region in the Fig. 4 corresponds to the region defined by the level of significance p of regarding an unauthorized person as the person specified by the ID. More specifically, the level of significance p of regarding an unauthorized person as the person specified by the ID is specified to determine that the speaker is the person specified by the ID when the distance d_{id} is in the range in which the level of significance of regarding an unauthorized person as the person specified by the ID is smaller than the designated level of significance p.

Next, Fig. 5 is a block diagram of a speaker verification apparatus of one example of the present invention when verifying the speaker. Referring to Fig. 5, numerals 51A and 51B denote DP matching parts. Numeral 52 denotes a statistic calculating part. Numeral 53 denotes a speaker judging part. Numeral 54 denotes a false acceptance error rate input part.

In Fig. 5, similarly to Fig. 3, an identity claim is input to the ID input part 31 at the time of using a system. Then, the speaker template selecting part 32 selects a template corresponding to the identity claim from templates of a plurality of speakers that are previously registered in the speaker template storing part 33 and sends the selected template to the DP matching part 51A. At the same time, the templates of the registered speakers other than the speaker corresponding to the identity claim are sent out to the DP matching part 51B. Herein, "DP" stands for dynamic programming.

Next, in the voice analyzing part 35, a voice input to the voice input part 34 is converted into a feature parameter for speaker verification and sent out to the DP matching calculating parts 51A and 51B. The DP matching part 51A calculates the distance d_{id} between the voice template of the speaker corresponding to the identity claim and the feature parameter of the input voice.

On the other hand, the DP matching part 51B calculates the distances d_1, d_2, ...and d_N between the voice templates of N other registered speakers and the feature parameter of the input voice, and delivers the results to the statistic calculating part 52. The statistic calculating part 52 estimates the average μ and the standard deviation σ of the speaker distances, using the calculated N distances d_1, d_2, ...and d_N with respect to the other registered speakers, and delivers the estimations to the speaker judging part 53. The speaker judging part 53 defines a normal distribution using the average μ and the standard deviation σ of the distances with respect to the other registered speakers.

If the probability distribution is a normal distribution, a probability distribution function F(d) in a point $\alpha \cdot \sigma$ away from the average μ can be determined by α. Therefore, whether or not the speaker is the person specified by the ID can be determined by examining whether or not the verification distance d_{id} is in a region where d_{id} is equal to or smaller than $(\mu - \alpha \cdot \sigma)$ in order to determine whether or not the verification distance d_{id} with resect to the input voice is within the region defined by the previously designated level of significance p of regarding an unauthorized person as the person specified by the ID. More specifically, $(\mu - \alpha \cdot \sigma)$ and d_{id} are compared and the determination is performed as follows. When d_{id} is equal to or smaller than $(\mu - \alpha \cdot \sigma)$, it is determined that the speaker is the person specified by the ID. When d_{id} is larger than $(\mu - \alpha \cdot \sigma)$, it is determined that the speaker is not the person specified by the ID. In the case where it is assumed that the probability distribution is a normal distribution, the false acceptance

error rate input part 54 inputs α corresponding to the level of significance p of regarding an unauthorized person as the person specified by the ID beforehand.

In this embodiment, the feature parameters are registered in the form of templates beforehand, and the probability distribution with respect to other registered speakers is estimated based on the speaker distances obtained by DP matching. The present invention is not limited to this method. For example, the probability distribution can be estimated based on a probability value output from a probability model such as Hidden Markov Model.

Furthermore, in the speaker template storing part 33, speakers may be classified by the gender beforehand. When the speaker corresponding to the identity claim is male, the speaker templates of other male speakers are used for estimation of the probability distribution. When the speaker corresponding to the identity claim is female, the speaker templates of other female speakers are used for estimation of the probability distribution. Thus, the error rate of the probability distribution becomes closer to the error rate obtained from the normal distribution function table. (The identity claim is something which indicates a specific individual such as a name).

Furthermore, in this embodiment, the probability distribution of the speaker distances is estimated as a single normal distribution. However, the probability distribution can be estimated as a mixed normal distribution defined by weighting addition of a plurality of normal distributions or other general probability distributions. (This is not necessarily limited to the distribution of other registered speakers, and other speakers can be prepared for the calculation of the distribution.)

Next, the effects of this embodiment are confirmed by the results of the following experiments. First, Fig. 6 is a graph showing the results of verification of 15 male speakers using the speaker verification method of this embodiment.

In Fig. 6, the horizontal axis indicates α obtained from the

11

normal distribution function according to the previously designated false

error rate.   The solid line indicates theoretical values of the false error

rate, which can be calculated as $1 - \Phi(\alpha)$ using a normal distribution

function $\Phi(\alpha)$ because the distribution of the speaker distances is

5   assumed to be a normal distribution.

Furthermore, FA (false acceptance) indicates a false acceptance

error rate, which is a probability of erroneously accepting an

unauthorized person.   FR (false rejection) indicates a false rejection

error rate, which is a probability of erroneously rejecting the person

10   specified by the ID.

In Fig. 6, the solid line shows the theoretical values of the false

acceptance error rate.   The short broken line shows FR obtained by

experiments, and the long broken line shows FA obtained by experiments.

As shown in Fig. 6, the solid line substantially matches the long broken

15   line, which means that the experimental results of the false acceptance

error rate are not significantly different from the theoretical values.

Therefore, the verification accuracy of the speaker verification method of

verifying the speaker based on the pre-assigned false acceptance error

rate of this embodiment is expected to be high.

20   Similarly to Fig. 6, Fig. 7 shows the verification results when a

white noise with a SNR (signal noise ratio) of about 20dB is added to the

voice to be verified.   Herein, "a SNR of about 20dB" refers to the level in

which noise is mixed in a ratio of one noise to 10 signals.   Furthermore,

the solid line shows the theoretical values of the false acceptance error

25   rate.   FR (noisy) indicates a FR when a white noise is mixed.   FR (clean)

indicates a FR when there is no white noise.   FA (noisy) indicates a FA

when a white noise is mixed.   FA (clean) indicates a FA when there is no

white noise.

The experimental results of Fig. 7 show that with respect to FR,

30   the white noise significantly changes the false rejection error rate, which

is the probability of erroneously rejecting the person specified by the ID.

On the other hand, with respect to FA, the white noise does not affect the

12

fact that the solid line for the theoretical values of the false acceptance error rate substantially matches the experimental results regarding FA. Therefore, even if an unexpected noise is input together with the voice because the voice input environment is varied, the verification accuracy of the speaker verification method of verifying the speaker based on the pre-assigned false acceptance error rate of this embodiment is expected to be high.

Furthermore, similarly to Fig. 6, Fig. 8 shows the verification results when the time gap between the input of a voice for verification and the input of voices for registration increases by 3 months. In Fig. 8, the solid line shows the theoretical values of the false acceptance error rate, and the experimental results of FA and FR after 3, 6, 9, and 12 months have passed are shown for each of FR and FA.

The experimental results of Fig. 8 show that with respect to FR, the time gap significantly changes the false rejection error rate, which is the probability of erroneously rejecting the person specified by the ID. On the other hand, with respect to FA, the time gap does not affect the fact that the solid line for the theoretical values of the false acceptance error rate substantially matches the broken lines indicating FA for every 3 months.

Therefore, even if the feature of the voice of the speaker has changed because of the time gap of the input of the voice, there is no significant change in the speaker distances with respect to other registered speakers. Thus, the verification accuracy of the speaker verification method of verifying the speaker based on the pre-assigned false acceptance error rate of this embodiment is maintained high. In addition, there is no need of updating the once-registered speaker templates every time the system is used, which eliminates an excessive burden on the user.

Next, Fig. 9 is a block diagram of a speaker verification apparatus of one example of the present invention when registering speakers. In Fig. 9, numeral 91 denotes a registration individual ID input part.

13

Numeral 92 denotes a registration voice input part.   Numeral 93 denotes a registration voice analyzing part.   Numeral 94 denotes a background noise input part.   Numeral 95 denotes a noise addition part.   Numeral 96 denotes a voice database regarding other registered speakers.

5   In Fig. 9, the individual ID of a speaker to be registered is input from the registration ID input part 91, and the voice of the speaker is input from the registration voice input part 92.   The voice input from the registration voice input part 92 is converted into a feature parameter in the registration voice analyzing part 93 and is stored in the speaker

10   template storing part 33 in linkage with the individual ID information as the voice template of the registered speaker.

Then, in order to match the input environment of the speaker to be registered to the voice database input environment of other registered speakers, background noise is input to the background noise input part 94.

15   Then, the noise addition part 95 adds the input background noise to the voice data of the other registered speakers in the voice database 96, which have been registered beforehand.   Herein, "background noise" refers to a noise that is inevitably input when a voice is input.   For actual input, only a noise that accompanies no voice before or after inputting a voice is

20   input.   Then, the registration voice analyzing part 93 converts the voice data with the noise into feature parameters in the same manner as the input voice corresponding to the individual ID.   Then, the speaker template storing part 33 stores the feature parameters as the voice templates of the other registered speakers at the same time when the

25   voice template of the registered speaker is stored.

This embodiment prevents the voice input environment of other registered speakers from being significantly different from the voice input environment of the speaker to be registered.   For example, even if the voice input environment of an unauthorized person is closer to the voice

30   input environment of the registered speaker than to the voice input environment of the other registered speakers, erroneous determination that the unauthorized person is the person specified by the ID can be

14

avoided.

In this case, the input environment is adjusted not necessarily only with respect to the voice data itself. The adjustment can be performed after the voice data are converted into a feature parameter of the voice. In addition, in the case where the voices of the other registered speakers are represented by probability models such as Hidden Markov Model, the environment can be adjusted by adapting registered speaker HMMs.

As described above, this embodiment where the speaker is verified based on the false acceptance error rate makes it possible to obtain the criterion for judging the speaker that is closer to the theoretical values of the statistical probability distribution and to maintain the false acceptance error rate closer to the theoretical values even if the voice input environment changes and a noise is mixed. Thus, the verification accuracy of the speaker verification can be maintained high without being affected by the aging of the input voice.

Next, the flow of processes of a program that realizes the speaker verification apparatus of an embodiment of the present invention will be described. Figs. 10 and 11 show flowcharts of processes of a program that realizes the speaker verification apparatus of an embodiment of the present invention.

First, Fig. 10 is a flowchart of processes for verifying a speaker in the speaker verification apparatus of an embodiment of the present invention. Referring to Fig. 10, a user inputs his/her individual ID and voice and a false acceptance error rate (step S101). The false acceptance error rate generally is previously input by a system manager as a predetermined value.

Then, the registered speaker corresponding to the individual ID is selected from the other registered speakers based on the individual ID (step S102). The data of the registered speaker corresponding to the individual ID are used to obtain the verification distance to the input voice, and the data of the other registered speakers are used to obtain the

15

probability distribution of the interspeaker distances.

Then, the feature parameter of the input voice is extracted (step S103), and the verification distance with respect to the registered speaker corresponding to the individual ID and the speaker distances with respect to the other registered speakers are calculated (step S104). The calculated results of the speaker distances with respect to the other registered speakers are used to estimate the probability distribution of the speaker distances (step S105).

When the probability distribution of the speaker distances is obtained, a region defined by the false acceptance error rate can be obtained in the probability distribution. Thus, it is determined whether or not the verification distance with respect to the registered speaker corresponding to the individual ID is included in the region (step S106). In the case where the verification distance with respect to the registered speaker corresponding to the individual ID is included in the region, the input voice is determined to be the voice of the registered person specified by the individual ID (step S107). In the case where the verification distance with respect to the registered speaker corresponding to the individual ID is not included in the region, the input voice is determined to be the voice of an unauthorized person (step S108).

Next, Fig. 11 is a flowchart of processes for registering a speaker in the speaker verification apparatus of an embodiment of the present invention. Referring to Fig. 11, a user inputs his/her individual ID and voice and background noise data (step S111).

Then, the voice data of the other registered speakers are obtained (step S112). The method of obtaining the voice data is not limited to a particular method, but it is preferable to prepare a database of the voice data regarding the other registered speakers beforehand.

Next, the input background noise is added to the obtained voice data of the other registered speakers (step S113). Thus, it is possible to minimize the difference between the environment of the input speech and that of other registered speakers beforehand in the input voice

16

environment.

Then, feature parameters are extracted with respect to the input voice and the voice data of the other registered speakers to which the noise is added (step S114). The feature parameter of the input voice

5 corresponding to the individual ID is stored as the speaker voice template. At the same time, the feature parameters of the voices of the other registered speakers are stored as the voice templates, which are used to calculate the speaker distances with respect to the other registered speakers (step S115).

10 A recording medium in which programs for realizing the speaker verification apparatus of the embodiment of the present invention are recorded can be not only a transportable recording medium 122 such as a CD-ROM 122-1, or a floppy disk 122-2, but also a remote accessible storage apparatus 121 or an equipped storage medium such as a hard

15 disk and a RAM of a computer, as shown in Fig. 12. The program 124 is loaded into the main memory of a data processing apparatus 123, and executed.

A recording medium in which the speaker templates or the like that are generated by the speaker verification apparatus of the

20 embodiment of the present invention are recorded can be not only a transportable recording medium 122 such as a CD-ROM 122-1, or a floppy disk 122-2, but also a remote accessible storage apparatus 121 or an equipped storage medium such as a hard disk and a RAM of a computer, as shown in Fig. 12. For example, the recording medium can be read by

25 a computer when using the speaker verification apparatus of the present invention.

The invention may be embodied in other forms without departing from the spirit or essential characteristics thereof. The embodiments disclosed in this application are to be considered in all respects as

30 illustrative and not limiting. The scope of the invention is indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the

17

claims are intended to be embraced therein.